IJDC | *Peer-Reviewed Paper*

The Digitized Archival Document Trustworthiness Scale

Devan Ray Donaldson
School of Informatics and Computing
Indiana University, Bloomington

Abstract

Designated communities are central to validation of preservation. If a designated community is able to understand and use information found within a digital repository, the assumption is that the information has been properly preserved. As judging the trustworthiness of information requires at least some level of understanding of that information, this paper presents results of a study aimed at developing a tool for measuring designated community members' perceptions of trustworthiness for preserved information found within a digital repository. The study focuses on genealogists at the Washington State Digital Archives who routinely interact with digitized genealogical records, including digitized marriage, death, and birth records. Results of the study include construction of an original Digitized Archival Document Trustworthiness Scale (DADTS). DADTS is a ready-made tool for digital curators to use to measure the trustworthiness perceptions of their designated community members. Implications of this study include the feasibility of engaging members of a designated community in the construction of a scale for measuring trustworthiness perception, thereby providing deeper insight into the understandability and usability of preserved information by that designated community.

Received 18 November 2015 ~ *Revision received* 26 August 2016 ~ *Accepted* 26 August 2016

Correspondence should be addressed to Devan Ray Donaldson, Ph.D., 1320 E. 10th St., Wells Library 019, Department of Information and Library Science, School of Informatics and Computing, Bloomington, IN, USA, 47405-3907. Email: drdonald@indiana.edu

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

For the past several years, digital curation researchers have focused on the concepts of a designated community and trustworthiness. Regarding the concept of designated communities, international standards require digital repositories to define and monitor their designated communities of users – the primary audience for the information they preserve (e.g., Deutsches Institut für Normung, 2012; International Organization for Standardization, 2012). These standards also require digital repositories to ensure that information preserved by repositories is understandable by their designated communities of users. However, methods of engaging designated community members to provide insight into their understanding of preserved information are lacking. Regarding the concept of trustworthiness, most research has focused on ways to ensure that digital repositories that are responsible for long-term preservation of information are actually up to the task (Dale and Gore, 2010; Ross and McHugh, 2005). Much less attention has been paid to understanding the trustworthiness of preserved information as perceived by the designated community members for whom that information is intended. More research could illuminate when and how digital repositories can play a role in helping designated community members judge the trustworthiness of preserved information.

A repository is trustworthy when its information is independently understandable to its designated community of users (Giaretta, 2011). At present, no measures exist for verifying the understandability of information a digital repository preserves. Since, to a certain extent, judging the trustworthiness of information involves inspection of content (Rieh, 2002; Wilson, 1983), this paper explores measurement of a designated community's perceptions of trustworthiness for content found within a digital repository as evidence in support of its understandability. The rationale for this idea is the assumption that it would be difficult for a user of content found within a digital repository to assess its trustworthiness if they can not understand it. In this respect, measuring designated community members' trustworthiness perceptions for preserved content provides some insight into its understandability, which, in turn, provides insight into whether a digital repository is successful at making information independently understandable and is therefore a trustworthy repository.

This study lies at the nexus of research on the concepts of a designated community and trustworthiness by examining a designated community's perceptions of trustworthiness for preserved information found within a digital repository. Specifically, the purpose of this study is to build, test, and assess a scale for measurement of preserved information trustworthiness perception. The remainder of this paper is as follows. First, the background section focuses on the concepts of a designated community and trustworthiness. Second, the methodology section describes the application of scale development, including how the data for the study were collected and analyzed. The methodology section also describes the digital repository that was used as the primary site of study as well as the designated community that was the focus of this study. Third, the findings section presents empirical, statistical results, which provide support and justification for an original Digitized Archival Document Trustworthiness Scale (DADTS). Fourth, the discussion section addresses how, under certain circumstances, digital curators can apply DADTS to understand their designated communities' perceptions of trustworthiness for preserved information. Finally, the

conclusion section examines the implications of this study for understanding designated communities' understanding and interpretation of preserved information and suggests directions for future research.

Background

Designated Communities

A designated community is a group of users that a digital repository identifies that should be able to understand the information[1] provided by that repository; the designated community may comprise multiple user communities (Consultative Committee for Space Data Systems, 2002). Designated communities are crucial for verifying the effectiveness of preservation efforts. According to Giaretta (2011), 'preserving digitally encoded information means that we must ensure that the information to be preserved is **Independently Understandable** to (and usable by) the Designated Community' (emphasis in original). In other words, one way to know if digital repositories are doing a good job at preservation is if the people who should be able to understand the preserved information can actually understand it. Given this, the question becomes, how to verify the understandability of preserved information by designated communities?

International standards for Trustworthy Digital Repositories (TDRs) take up discussion of verifying the understandability of preserved information by designated communities. For example, ISO 16363 requires use of 'test procedures' that ensure the understandability of preserved information by designated communities (International Organization for Standardization, 2012). As another example, DIN 31644 requires digital repositories to 'check at regular intervals' whether preserved information can still be interpreted by the designated community or communities (NESTOR Certification Working Group, 2013). In both cases, however, the standards currently do not include recommendations or describe specific ways of engaging designated community members that would in any way demonstrate their understanding of preserved information. Consequently, this offers an opportunity for research on identifying approaches to verifying understandability.

Recent research on designated communities has focused on ways of identifying them and monitoring their knowledge bases over time through analysis of their use of websites and social media (Kim, 2015), and has also focused on their perceptions of trust in digital repositories (Yakel, Faniel, Kriesberg, and Yoon, 2013). Taken together, recent research on designated communities demonstrates that we can devise a way for a repository to identify a designated community and we can understand the extent to which it perceives a repository as trustworthy. Questions remain about how to verify the extent to which designated community members understand information preserved by digital repositories, as their understanding serves as a measure of preservation effectiveness.

1 This paper borrows Buckland's (1991) use of the term information. According to Buckland, 'the term "information" is also used attributively for objects, such as data and documents, that are referred to as "information" because they are regarded as being informative.'

Trustworthiness

There is no universally agreed upon definition of trustworthiness. The concept can mean different things in different contexts; it depends upon who is judging it and at what level it is being considered. For example, at the repository level, trustworthiness typically refers to a repository's ability to preserve and provide access to digital content over the long term (Ross and McHugh, 2005). In more traditional environments, such as libraries, trustworthiness can refer to honesty and a lack of deception (Wilson, 1983). In archives, whether traditional or digital, it seems that one way to define trustworthiness at the document level[2] is in terms of authenticity (e.g., a document has not been altered or changed since its original creation) and reliability (e.g., the degree to which the document accurately reflects what happened) (Duff, Cherry, and Craig, 2004; Duranti, 1995; MacNeil, 2000).

Although definitions of information trustworthiness found in online environments (e.g., the Internet and digital repositories accessible online) vary in some respects, all seem to acknowledge the necessity of user judgments. Examples of these definitions include 'when it appears to be reliable, unbiased, and fair' (Hilligoss and Rieh, 2008), and 'a receiver judgment based primarily on subjective factors' (Flanagin and Metzger, 2008). Of the studies that explore the trustworthiness of information found in digital repositories, not all researchers define the term. For example, neither Van House (2002) nor St Jean, Rieh, Yakel, and Markey (2011) explicitly define trustworthiness at the document level.

People employ systematic and/or heuristic processing to judge the trustworthiness of online information (Sundar, 2008). Systematic processing involves assessment of content to arrive at trustworthiness judgments. In contrast, heuristic processing does not involve assessment of content, but involves reliance on cues related to information objects to make trustworthiness judgments. Studies have shown that people tend to employ heuristic evaluation rather than systematic evaluation of content because the latter requires more cognitive effort; this is particularly the case for novices, or people who are not domain experts or experienced researchers (Chaiken, 1980; Metzger, 2007). People employ a systematic evaluation of information when they believe that they possess domain expertise and they have the motivation to do it (Metzger, 2007). In practice, people often rely on heuristic processing, but also rely upon systematic processing or some combination of both when judging trustworthiness.

People rely on cognitive authorities, who they deem appropriate sources, to judge the trustworthiness of information. Wilson (1983) specifies four tests for recognizing the cognitive authority of information: 1) the cognitive authority of its author, 2) publication history, 3) authority of text type, and 4) intrinsic plausibility. Regarding the first test of cognitive authority, 'we can trust a text if it is the work of an individual or group of individuals whom we can trust' (Wilson, 1983). Informed decisions about the trustworthiness of an author or group of authors come from their reputations. Regarding the second test of cognitive authority, 'a publication house can acquire a kind of cognitive authority – not that the house itself knows anything, but that it is thought to be good at finding those who do and publishing their work' (Wilson, 1983). For example, the Oxford University Press has earned a reputation for publishing high-quality work, and thus acts as a cognitive authority. We trust books published by the Oxford University Press because we trust that they only publish trustworthy work. Regarding

2 This paper refers to data and documents interchangeably because I borrow Buckland's (1991) definition of information as thing, which casts data and documents similarly because they have the common feature of being informative.

the third test of cognitive authority, certain texts possess cognitive authority because of their type, not because of who wrote them. Examples of these include dictionaries and encyclopaedias. Finally, the fourth test of cognitive authority involves evaluation of the content itself; 'a text usually has only one chance to capture our attention and interest; reading a few words of it may be enough to discourage us from continuing or may lure us on to reading the whole thing' (Wilson, 1983). We trust the information because it looks and seems plausible, given what we know.

On the Internet, people rely on one or a combination of cognitive authorities to judge the trustworthiness of information. For example, Rieh (2002) found that people rely on characteristics of sources of information, including author/creator credentials, author/creator reputation, type of source, and URL domain type (e.g., .edu, .org, .com, .etc) to judge the trustworthiness of online information. In addition, Rieh (2002) found that people rely on characteristics of information objects, including content, type of information object, title, organization/structure, and presentation to judge the trustworthiness of online information.

Multiple studies of trustworthiness for information found within various types of digital repositories, including digital archives, digital libraries, institutional repositories, and domain-specific repositories, underscore the importance of similar kinds of cognitive authorities. For example, St Jean et al. (2011), Van House (2002), Van House (2003), and Van House, Butler, and Schiff (1998) found that the reputation of the author affects people's perceptions of trustworthiness for information they encounter in a digital repository. Also, if a designated community member perceives that the author has a good reputation, they will trust that information. People also assess the content when judging the trustworthiness of information they find in a digital repository. For example, in Fear and Donaldson (2012), participants reported that they would run tests on data and compare those results with what would seem reasonable and appropriate, given the parameters of the research instruments and data, in order to arrive at trustworthiness judgments for data that they themselves did not create.

Empirical studies exploring the effect of repository trustworthiness perception on document trustworthiness perception have had varying results. Multiple studies have found that trustworthiness perception at the repository level affects trustworthiness at the document level. For example, Van House (2003) found that designated community members often take for granted the trustworthiness of a digital library based on their knowledge of the institution; the digital library is assumed to adhere to some collection development standards and procedures that, to some degree, warrant its contents. In this respect, at least some of trustworthiness at the repository level trickles down to the resources within the repository thereby affecting perceptions of trustworthiness for content. Similarly, Fear and Donaldson (2012) found that the mere presence of a dataset within a digital repository provided at least some evidence of its trustworthiness. The participants did not think a researcher would take the effort to make the data available if the data were not trustworthy enough. St Jean et al. (2011) found that a digital repository's tie with an institution, e.g. a college or university, positively influenced their study participants' perceptions of the content that they found within the digital repository. The participants did not think the institution would allow low-quality content in the digital repository because it would compromise the institution's reputation. In contrast to these studies, Yakel et al. (2013) found no relationship between trust in digital repositories and trust in data; it was possible for the designated community members who participated in their study to trust the repository without necessarily trusting data found within that repository.

In summary, existing research on how trustworthiness operates is particularly relevant for motivating the current study. Although trustworthiness at the repository level is distinct and different from trustworthiness at the document or content level, depending on how designated community members choose to define trustworthiness at both levels, and depending on what factors they allow to influence their perceptions of trustworthiness at both levels, perception of trustworthiness at one level can affect perception of trustworthiness at another level. For example, perception of trustworthiness at the repository level could affect perception of trustworthiness at the document level. In addition, if in fact members of a designated community employ systematic evaluation of content to judge its trustworthiness, or they combine systematic and heuristic evaluation to arrive at trustworthiness judgments, evaluation of their trustworthiness perception could serve as an indication of the understandability of that information. The premise is that designated community members would have to understand the content well enough to conduct systematic evaluation and subsequently judge its trustworthiness. In other words, focusing on designated community members' trustworthiness judgments could be valuable for verifying that a digital repository is doing a good job at information preservation.

As an initial step toward understanding the relationship between repository and document trustworthiness perception, and as an initial step toward understanding the relationship between measurement of document trustworthiness perception and the understandability of preserved information, the current study centers on the following research question:

- To what extent are designated community members' perceptions of document trustworthiness measurable?

A scale for measurement of document trustworthiness perception could be compared with measurement of repository trustworthiness perception to examine the impact of one level of trustworthiness perception on the other. However, first we must investigate whether it is possible to construct a scale for measurement of document trustworthiness perception. Also, the measurement of document trustworthiness perception could serve as evidence of the understandability of that information, if, for example, systematic or heuristic processing is part of designated community members' document trustworthiness perception judgment process. However, before drawing any conclusions about the understandability of preserved information based on measurement of designated community members' perceived trustworthiness of that information, we must first investigate whether it is possible to construct a scale for measurement of document trustworthiness perception.

Methodology

To more deeply understand the interplay among the concepts of a designated community, the understandability of information preserved by a digital repository, and the trustworthiness of that information, this study employs the methodology of scale development (DeVellis, 2012). Researchers have employed similar methods to develop scales for measuring the perceived trustworthiness of people (e.g., Rotter, 1967; Rotter, 1971), online vendors (e.g., Gefen, 2002), and information found on the Internet (e.g., Fogg et al., 2001). This study is unique in that it applies scale development to the

concept of trustworthiness from the perspective of actual designated community members who frequently utilize information preserved by a digital repository.

Primary Site of Study and Designated Community

The Washington State Digital Archives (WADA) was the primary site of study. It was chosen for three primary reasons. First, WADA is a heavily utilized digital cultural heritage resource, developed and maintained at taxpayer expense as a mechanism for providing open, public access to archives and records of the State of Washington. Approximately 500,000 people visit the home page of WADA per year, with thousands of unique visitors per month. Second, WADA has a strong and explicit mission statement that focuses on making preserved digital information accessible to users (Washington State Archives – Digital Archives, 2016). Third, WADA conforms in principle to the requirements of a Trustworthy Digital Repository (TDR). It abides by leading best practices and standards for organizational infrastructure, digital object management, and technical infrastructure, including security issues, consistent with the International Organization for Standardization's specifications, despite not being formally certified as a TDR as of the time of this study (T.S. Badger, personal communication, March 8, 2013).

Study participants included experienced genealogists because this population of users represents WADA's largest designated community (T.S. Badger, personal communication, March 8, 2013). Also, based on WADA's statistics, genealogical records are among WADA's most highly downloaded documents. These designated community members typically download digitized marriage, death, and birth records available in JPEG format for use.

Scale Development

Scale development involves four primary steps (DeVellis, 2012; Spector, 1992):

- Step 1 – Construct Definition
- Step 2 – Generate an Item Pool
- Step 3 – Design the Scale
- Step 4 – Full Administration and Item Analysis.

Step 1 of scale development is to construct a definition. In this study, this involved a review of the literature to identify the scope of trustworthiness for the purpose of empirical investigation. Step 1 also involved focus groups to understand how members of a designated community (i.e., genealogists) talk about trustworthiness. The findings from the focus groups are reported elsewhere (Donaldson and Conway, 2015). Step 2 of scale development is to generate an item pool. This involved identifying items for measuring trustworthiness from multiple sources, including the literature, subject matter experts, and focus groups data (Donaldson and Conway, 2015). Step 3 of scale development is to design the scale. This involved transforming the item pool resulting from Step 2 into a web survey for pretesting and refinement. Step 4 of scale development is the full administration of the survey and subsequent item analysis. This involved administering the final item pool comprising items gathered from earlier steps of scale development to a large sample of designated community members for their evaluation. Each item described a circumstance one might encounter while using a

digitized archival document. Participants answered whether the circumstance described by each item would cause them to perceive a digitized archival document as either untrustworthy or trustworthy on a seven-point scale ranging from 'very untrustworthy' to 'very trustworthy.' An eighth option, 'Not Applicable,' was included for participants to choose if the circumstance an item described was not relevant to their experience of using digitized archival documents. Step 4 also involved analyzing designated community members' responses via factor analysis to identify the items that were most essential for measuring the trustworthiness of preserved information (in this case, digitized genealogical records).

Participant Recruitment

During Step 4, the survey instrument resulting from Step 3 was administered to a large sample of participants via Qualtrics online survey software. Participants were recruited via the intercept survey method (Couper, 2000). WADA staff set their web site to provide a pop up invitation with a link to the survey to between 13% - 30% of visitors to WADA's homepage. To increase participation, two data collection periods were held: one in December 2013 and one in February 2014. Only the responses of participants who provided a response for every item were analyzed.

Data Analysis

Two types of analysis were performed: item analysis and exploratory factor analysis. Item analysis involved analysis of item variances, item total correlations, item means, and item standard deviations (DeVellis, 2012). To assess item variances, the range of responses (i.e., items' minimums and maximums) for each item were inspected. To assess item-total correlations, each item was examined to determine the extent to which it correlated with the collection of remaining items. Items' means and standard deviations were examined to ensure that, for each item, the means were near the midpoint of the seven-point scale on which participants rated the items while also ensuring that there was variation involved in attaining the means.

After performing item analysis, exploratory factor analysis (EFA) was conducted using SPSS Statistics 22.0, a software package for statistical analysis, to establish the factor structure of the trustworthiness items (Kline, 2013). EFA was used as a tool to help identify latent variables that may explain correlation in the variables proposed as indicators of trustworthiness. 'Important' trustworthiness items were operationalized as items with high factor loadings on factors with large eigenvalues.[3] To assign items to factors, factor loadings equal to or higher than .32 were considered (Tabachnick and Fidell, 2001). Two tests were performed to assess the appropriateness of the data that were collected during this study for EFA: the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (Kaiser, 1960; Kaiser and Rice, 1974) and Bartlett's Test of Sphericity (Bartlett, 1954). Afterwards, EFA was conducted using principal axis factoring with oblique rotation (Kline, 2013). Results of Cattell's (1966) scree test were used to

3 In factor analysis, it is assumed that the reason items correlate is because they are related to the same concept (i.e., they share a common factor). Factor loadings demonstrate the degree to which items correlate with factors. Eigenvalues are statistical measures of how much variance factors explain. The greater the eigenvalue, the more the factor reflects the concept under investigation, in this case, trustworthiness. Items with high factor loadings are the result of factors that contribute greatly to those items. For these reasons, the results focused on items with high factor loadings on factors with large eigenvalues. For more on exploratory factor analysis, see Kline (2013).

determine the number of factors to retain. Donaldson (2015) contains the raw data set, the processed/analysed data in two different file formats (e.g., .doc and .spv), and the syntax file that lists the code that was used to perform the analyses.

Reliability and validity

The calculation of Cronbach's alpha was used to assess the reliability of the scale that was identified as a result of EFA (Cronbach, 1951). Content validity was accounted for by deriving items for the study from three sources: 1) the responses of focus group members of the same designated community regarding their perceptions of trustworthiness for the same types of digitized archival documents that were considered in this study (Donaldson and Conway, 2015), 2) the literature on document trustworthiness, and 3) the recommendations of researchers who have studied trustworthiness. Regarding the first source, in prior research, I asked a different sample of genealogists who utilize WADA to describe what makes a digitized document trustworthy in their opinion (Donaldson and Conway, 2015). I turned their responses into survey items that I used for this study. For example, during the focus groups, when I asked what makes a document trustworthy, one of the participants responded, 'Is it factual?' implying that if a document is factual, then it is trustworthy. I turned that response into the survey item 'The document is factual' for the purpose of this study. Regarding the second source, I surveyed the literature on trustworthiness, including trustworthiness perception measurement. For example, Duff et al. (2004) provided the definition 'the degree to which the record accurately reflects what happened' before asking their participants a question about the perceived reliability of archival materials. They asked this question as part of measuring archival trustworthiness perception. I transformed their definition into the survey item 'The document accurately reflects what happened' for the purpose of this study. Regarding the third source, I asked researchers who have conducted research on trustworthiness to recommend items for trustworthiness perception measurement. They provided me with lists of potential items and I selected items from those lists for inclusion in this study.

Findings

Following the steps of scale development produced a pool of trustworthiness items and a dataset which included a designated community's evaluation of those items. This section reports on analysis of this dataset.

It was clear, based upon the sources of the items in this study (e.g., the literature, trustworthiness subject matter experts, and actual designated community members' responses regarding the concept of trustworthiness during focus groups), that the items were related to trustworthiness. However, it was not clear exactly *how* the items related or which items were most important for measurement. Using the designated community members' ratings of each of the trustworthiness items, EFA was performed to establish the relationship among the trustworthiness items and identify which items were most important for measurement of trustworthiness.

The study participants evaluated 74 items that were related to some aspect of the trustworthiness of preserved information (in this case, digitized genealogical records). 48 of those items were omitted from further analysis because ten or more participants indicated that those items were not applicable to their concept of trustworthiness. After

discussing participant characteristics, this section reports on the results of EFA for the remaining 26 trustworthiness items.

Participant Characteristics

Since the main goal of the study was to develop a measure of trustworthiness perception for preserved information, an appropriate sample had to include members of a specific designated community who frequently utilize specific types of preserved information. Prior contact with the Deputy State Archivist of the Washington State Archives confirmed that WADA's largest designated community of users included genealogists (T.S. Badger, personal communication, March 8, 2013). Thus, an appropriate and representative sample for this study required participants with demographics similar to the larger population of genealogists as well as experience in using preserved information, in this case, digitized archival documents.

While there is no sampling frame available to assess genealogists as a population, there are multiple studies of genealogists in archival science and information science. Most of what is known about genealogists in this literature is based upon samples of predominantly older females (Case, 2008; Yakel, 2004). A profile of demographic characteristics was derived from these studies and used as a proxy for the demographics of the larger population. This study population was comprised almost completely of older adults; 96% (n=172) reported that they were 40 years of age or older. The remaining seven participants were either between 30 and 39 (n=3) or 20 and 29 (n=4). The study population was also predominantly female; almost three quarters (74%) of the participants reported that they were female. The remaining 26% reported that they were male.

In addition to demographic characteristics, this study also required that participants have experience using information preserved by a digital repository. Participants answered questions related to their primary reason for using WADA documents, frequency of using WADA documents, frequency of use by document type, and time spent using documents on a typical visit to WADA. All study participants reported that their primary reason for using WADA documents was to conduct genealogical research. Nearly three quarters of the survey sample reported that they used WADA documents daily (4%), weekly (28%), or monthly (42%). Nearly a quarter (24%) of the participants indicated that they used WADA documents a few times a year. The remaining four participants indicated that they had not used WADA documents within the last year. The study participants reported using digitized marriage records most frequently (55%) followed by death records (31%), census records (3%), birth records (2%), and land records (1%). Fourteen participants (8%) used the 'other' category to indicate that they used birth, death, marriage, census, and land records. Nearly half (47%) of the participants indicated that they spent either over 30 minutes to an hour (29%) or between one and two hours (18%) using WADA documents on a typical visit. Ten percent reported that they spent over two hours using WADA documents per visit. The remaining 43% indicated that they spent between 0 to 30 minutes using WADA documents on a typical visit.

Overall, these findings suggest that it was possible for the study participants to evaluate the trustworthiness items in this study based on the quantity and quality of their reported experiences with WADA documents. This was critical for this study, which focuses on trustworthiness perception regarding information preserved by a digital repository. Nunnally and Bernstein (1994) recommend participation of four to ten

subjects per item in a scale development project. Since 26 items were used during the EFA, and 179 participated in this study, the sample size is sufficient.

Exploratory Factor Analysis Results

The following discusses results of appropriateness of the data for EFA, describes results of the scree test, and reports on the items that were associated with the factor that was retained as a result of the data analysis.

Results of tests of appropriateness of the data for EFA

Results of two tests verified the appropriateness of the data for use of EFA: the Kaiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett's Test of Sphericity.The results showed that the Kaiser-Meyer-Olkin index was .95, far exceeding the recommended value of .60 (Kaiser, 1960), whereas Bartlett's (1954) sphericity test was significant at a level of .000, ensuring that the data have sufficient correlations and justifying the use of EFA.

Scree test results

Results of one test verified the factor structure underlying the trustworthiness items. Cattell's (1966) scree test clearly showed a single-factor structure (F1, Eigenvalue = 13.76), which accounted for 51.4% of the total variance (see Figure 1). Relative to the first factor, none of the other factors held enough explanatory power. As shown in Figure 1, the second and all subsequent factors have eigenvalues of about one. This means that the explanatory value of any of those factors is not much more than the explanatory power of any particular trustworthiness item associated with those factors. Since the goal of EFA is a parsimonious account of factors and items, the goal of analysis is to identify the structure that explains the most variance using the fewest factors (DeVellis, 2012). Consequently, these results confirmed that use of only one factor could provide a parsimonious yet informative account of trustworthiness. As a result, only the first factor was retained for further analysis. Thus, there is strong empirical support for only one factor – the trustworthiness factor.

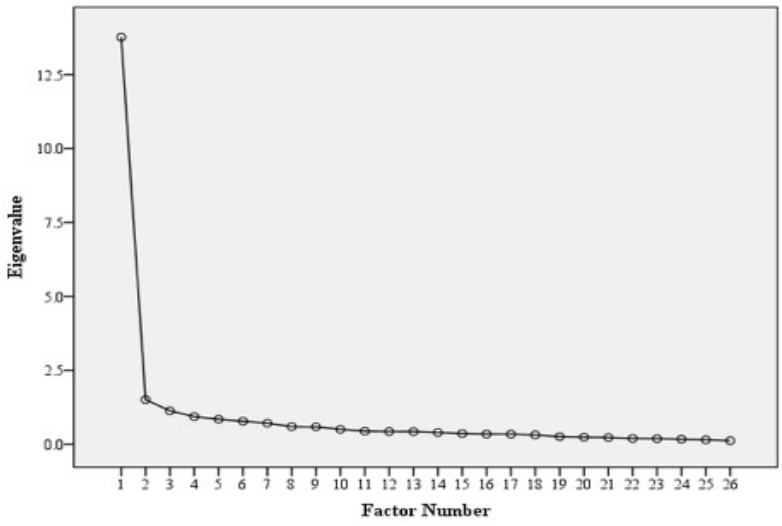

Figure 1. Scree plot.

Identification of most important trustworthiness measurement items

Only 12 of the 26 items that were considered during the EFA loaded strongly onto the only factor that was retained for further analysis – the trustworthiness factor. Table 1 lists those items. All of those items presented high factor loadings, exceeding the .32 cut-off recommended by Tabachnick and Fidell (2001). According to DeVellis (2012), items that load strongly onto a factor represent items as a scale for measurement of that factor. Hence, the items in Table 1 represent items in a scale for measurement of trustworthiness regarding digitized archival documents – The Digitized Archival Document Trustworthiness Scale (DADTS).

Results of one test confirmed the reliability of DADTS. Cronbach's alpha for the scale was .93 (Cronbach, 1951), which is well above the benchmark of .70 that is common in the social sciences (Hair, Black, Babin, and Anderson, 2010).

Table 1. The Digitized Archival Document Trustworthiness Scale (DADTS).

Item Number	Item	Factor Loading[4]
1	The document is authentic.	.93
2	The document is factual.	.93
3	The document includes documentation of where it came from.	.77
4	The document was created using responsible and accepted practices.	.68
5	The digitized document is an actual picture of the original physical document.	.68
6	The document is credible.	.66
7	The document appears free from error.	.59
8	The document is what it claims to be.	.54
9	The document is a primary source.	.54
10	The document accurately reflects what happened.	.45
11	The document is official.	.44
12	The document was written at the time of the event.	.37

Discussion

For the designated community members who participated in this study, the items comprising the Digitized Archival Document Trustworthiness Scale (DADTS) best illustrate their concept of trustworthiness for the types of preserved information they frequently utilize within the WADA digital repository. The participants evaluated 74 trustworthiness items. They rated each item on a seven-point scale ranging from -3, very untrustworthy, to +3, very trustworthy. The scale also included an eighth option, 'not applicable,' for participants to choose if they felt the circumstance described in any particular item was not indicative of their concept of trustworthiness. The participants could have decided that any of the 12 items comprising DADTS were not applicable to

4 The factor loadings presented in this paper are pattern coefficients. Pattern coefficients represent the unique contribution of each factor to each item (Hair et al., 2010). Items with large pattern coefficients are items that factors contribute the most to and are thus most indicative of those factors. In this case, these items are most indicative of trustworthiness.

their concept of trustworthiness by choosing the 'Not Applicable' rating. Instead, none of the participants chose this option; they associated the DADTS items with high and positive perceptions of trustworthiness for the types of preserved information they frequently encounter while using WADA.

Also, data analysis revealed that the 12 items comprising DADTS (see Table 1) were highly correlated with one another and the 12 DADTS items loaded strongly onto the only factor explaining a significant amount of variance. These findings demonstrate that only the 12 DADTS items stand out as being the most important indicators of trustworthiness. In other words, the participants' evaluation of the 12 items comprising DADTS yields more information about their concept of trustworthiness than any of the other items the participants rated during the study. Overall, a strength of this study is that participants had the power to choose what related and what was not relevant to their concept of trustworthiness. As a result, the findings demonstrate what really matters to this designated community with respect to their trustworthiness perceptions for preserved information, in this case digitized genealogical records.

48 items were omitted from the study because ten or more participants rated these items as 'Not Applicable' to their concept of trustworthiness. Examples of these items include: 'The person the document is about was alive during the time the document is created,' 'The document is handwritten,' 'The document appears blurry,' 'Names are misspelled,' and 'The document is typed.' It is not surprising that participants disagreed on the importance of these and similar items during the study. For example, just because a document is typed or handwritten does not necessarily mean it is trustworthy, although some people use information about how a document is written to inform their trustworthiness judgments. Notwithstanding, I used these items in this study to explore the extent to which they were applicable to larger sample of genealogists. Results of this study suggest that those items were not as applicable as the items that were retained for further analysis.

26 items were included for analysis in this study, including EFA, while only 12 of those items comprise DADTS. Examples of the 14 items that were not included in DADTS are: 'The document is legible,' 'The document includes verifiable data,' and 'The document is from the time period it claims to be.' These items were omitted from DADTS because they did not correlate as strongly with the other trustworthiness items as those that were included in DADTS. The weaker correlations of those 14 items suggest that they are not as good of trustworthiness indicators as the 12 DADTS items. This does not mean that the 14 items that were not included in DADTS are not relevant to the participants' concept of trustworthiness. It simply means that, relative to the items that were actually included in DADTS, they are not as important for measuring trustworthiness.

The results of this study suggest that users use multiple variables in combination to assess trustworthiness. Items that serve as indicators of trustworthiness represent multiple different concepts, including authenticity, provenance, reliability, credibility, and accuracy. For example, the item 'The document is what it claims to be' represents the concept of authenticity, the item 'The document includes documentation of where it came from' represents the concept of provenance, the item 'The document accurately reflects what happened' represents the concept of reliability, the item 'The document is credible' represents the concept of credibility, and the item 'The document appears free from error' represents the concept of accuracy. Although researchers in multiple fields, such as digital curation and web credibility, have drawn distinctions among these concepts, there was not empirical support for dividing the items that represent these different concepts into separate subscales for measurement of trustworthiness

perception. Instead, there was strong empirical support for including the 12 items pertaining to the concepts of authenticity, provenance, reliability, credibility, and accuracy together in a single scale for measurement of trustworthiness perception.

In addition, the findings underscore the importance of what Wilson (1983) calls the intrinsic plausibility aspect of cognitive authority in judging the trustworthiness of preserved information found within a digital repository. The intrinsic plausibility test of cognitive authority pertains to the first impression a text makes upon one's visual inspection of it. Ten out of the 12 items (items 1, 2 and items 5 through 12) comprising the Digitized Archival Document Trustworthiness Scale (DADTS) correspond to some aspect of the content which participants evaluated in relation to their trustworthiness perceptions. In particular, multiple DADTS items pertain to the concept of authenticity (e.g., 'The document is authentic,' 'The document is what it claims to be') underscoring the importance of the concept of authenticity to the participants' concept of trustworthiness. As well, items pertaining to reliability (e.g., 'The document accurately reflects what happened') underscore the importance of this concept to the participants' concept of trustworthiness. The fact that inspection of the cognitive authority of text via intrinsic plausibility as well as the concepts of authenticity and reliability all relate to the participants' concept of trustworthiness is not new, as these findings are consistent with prior theories of and research on the concept. What is new is DADTS offers a means of measuring the impact of these concepts with regard to a designated community's perception of trustworthiness for preserved information found within a digital repository. DADTS provides an empirical, statistical measure of the influence of cognitive authority, authenticity, and reliability perception on a specific designated community's concept of trustworthiness.

The findings also demonstrate the importance of the cognitive authority of the author or creator in the designated community's judgment of the trustworthiness of information preserved by a digital repository. Multiple DADTS items relate to the author or creator of the information (e.g., 'The document includes documentation of where it came from,' 'The document was created using responsible and accepted practices'). The influence of the cognitive authority of the author or creator of information on trustworthiness perception is not new. What is new is DADTS offers a means of measuring the impact of the cognitive authority of the author or creator of the information with respect to the participants' concept of trustworthiness. Ultimately, DADTS brings together in one scale tests of cognitive authority which focus on inspection of content and also takes into account the cognitive authority of the author or creator of the information. Consequently, DADTS is actually sensitive to the nuances of the designated community members' trustworthiness perceptions.

No bona fide measures of the understandability of preserved information by a designated community exist. DADTS represents an initial step in that direction. Most DADTS items correspond to assessment of content, suggesting that the designated community members who participated in this study must have had at least some level of understanding of the preserved information in order to provide ratings for the DADTS items. Going forward, digital curators with similar designated communities and similar collections can administer DADTS items to their users as a means of assessing the extent to which they understand information preserved by their digital repositories. Specifically, digital curators can use their designated community members' ratings of DADTS items as evidence of perceived understandability thereby addressing criteria in standards for Trustworthy Digital Repositories related to understanding and monitoring designated communities as well as ensuring the understandability and usability of preserved information by designated communities.

Some digital curation researchers have argued that there is a relationship between repository and document trustworthiness. Others have argued that there is no such relationship. While repository trustworthiness and document trustworthiness are separate and distinct phenomena, there could be a relationship between the two. Definitions and measures of trustworthiness at both the repository level and the document or content level would provide a means of addressing this question empirically. At present, metrics exist for measurement of trustworthiness at the repository level (Dale and Gore, 2010). Standards for certification of Trustworthy Digital Repositories (TDRs), such as the Data Seal of Approval, the World Data System certification program, ISO 16363, and DIN 31644, represent these metrics. Within the context of these standards, repository trustworthiness refers to demonstration of a repository's ability to preserve digital information for the long term (Giaretta, 2011; Ross and McHugh, 2005). Recent studies have begun to examine the concept of repository trustworthiness from the point of view of members of a designated community – those who should be able to understand the information preserved by a digital repository (e.g., Yakel, Faniel, Kriesberg, and Yoon, 2013).

At the document or content level, few measures of trustworthiness exist. DADTS offers a user-oriented approach to measurement of trustworthiness at the document level. Digital curators could use DADTS to measure user document trustworthiness perception and compare those measurements against measures of repository trustworthiness. For example, digital curators could measure user perceptions of document trustworthiness within a repository that has been formally certified as trustworthy and also measure user perceptions of document trustworthiness within a repository that has not been formally certified as trustworthy for purposes of comparison. One could hypothesize that the repository that has been certified as trustworthy will receive higher user document trustworthiness perception ratings than the repository without trustworthy certification. At any rate, DADTS is an example of a tool that would allow for this type of empirical statistical comparison across different repositories. In addition, DADTS, which focuses on document trustworthiness perception within a digital repository context, could be combined with repository trustworthiness perception measures to empirically examine the effect of document trustworthiness perception on repository trustworthiness perception, and vice versa.

There is one primary limitation of the research; the generalizability of the findings has not yet been examined empirically. If a different sample were given the same set of 26 items to evaluate, the 12 items comprising DADTS may or may not be the same items that load onto the trustworthiness factor after factor analysis. Future studies could administer the same set of items used in this study to a different sample of designated community members, perform EFA, and compare the results of that study to the results of this study to determine whether DADTS emerges consistently as a finding across different samples.

Conclusion

Digital curation has become established as a distinctive domain of professional practice, bounded by a suite of international standards, determined by an international network of best practices, and founded on the principle that long term preservation depends on the development and persistence of trustworthiness. The emergence of Trustworthy Digital Repositories as viable storehouses of data, information, and knowledge is the most

compelling evidence to date that digital curation practices are capable of affecting the long term preservation of digital information. Digital repositories are special domains of managed information. By intention and design, repositories protect the authenticity and accuracy of digital documents, and in doing so, establish and maintain their trustworthiness. The results of this study demonstrate that it is possible and valuable to measure with statistical soundness and conceptual nuance how a designated community of users perceives the trustworthiness of digital information. In doing so, the original Digitized Archival Document Trustworthiness Scale (DADTS) presented in this paper lays the groundwork for future investigations of how trustworthiness, beyond serving as a symbolic brand, truly functions as an operational component of digital repositories.

Acknowledgements

The author would like to thank Paul Conway, William Jacoby, James Lepkowski, Soo Young Rieh, and Elizabeth Yakel for their guidance on the development of this project. The author would also like to thank the Horace H. Rackham School of Graduate Studies at the University of Michigan for its financial support of this project. The author would also like to thank Brady West and colleagues at Consulting for Statistics, Computing and Analytics Research (CSCAR) for their assistance with data analysis. The author would also like to thank Howard Rosenbaum, Ronald Day, Charles Senteio, Lois Street, Delano Small, and Ryan Davidson for their feedback on earlier drafts of this paper.

References

Bartlett, M.S. (1954). A note on the multiplying factors for various χ2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *16*(2), 296-298.

Buckland, M.K. (1991). Information as thing. *Journal of the American Society for Information Science*, *42*(5), 351-360.

Case, D.O. (2008). Collection of family health histories: The link between genealogy and public health. *Journal of the American Society for Information Science and Technology*, 59(14), 2312-2319.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1(*2), 245-276. doi:10.1207/s15327906mbr0102_10

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, *39(*5), 752-766. Retrieved from http://www.unc.edu/~fbaum/teaching/articles/jpsp-1980-Chaiken.pdf

Consultative Committee for Space Data Systems. (2002). Reference model for an open archival information system (OAIS) recommendation for space data system standards. Blue Book. Washington, DC: CCSDS Secretariat. Retrieved from http://dhcommons.tamu.edu/sites/default/files/ccsds_oais_model.pdf

Couper, M.P. (2000). Review: Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*(4), 464−494. doi:10.1086/318641

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. Retrieved from http://psych.colorado.edu/~carey/Courses/PSYC5112/Readings/alpha_Cronbach.pdf

Dale, R.L., & Gore, E.B. (2010). Process models and the development of trustworthy digital repositories. *Information Standards Quarterly*, *22*(2), 14-19.

Deutsches Institut für Normung. (2012). *Information and documentation – criteria for trustworthy digital archives.* Berlin: Deutsches Institut für Normung.

DeVellis, R.F. (2012). *Scale development: Theory and applications*. Thousand Oaks, California: SAGE.

Donaldson, D.R. (2015). *The digitized archival document trustworthiness scale study dataset and associated files* [Data set]. Bloomington, IN: IU ScholarWorks. Retrieved from http://hdl.handle.net/2022/20510

Donaldson, D.R., & Conway, P. (2015). User conceptions of trustworthiness for digital archival documents. *Journal of the Association for Information Science and Technology*, *66*(12), 2427-2444. doi:10.1002/asi.23330

Duff, W., Craig, B., & Cherry, J. (2004). Finding and using archival resources: A cross−Canada survey of historians studying Canadian history. *Archivaria*, *58*, 51-80. Retrieved from http://archivaria.ca/index.php/archivaria/article/view/12478/13590

Duranti, L. (1995). Reliability and authenticity: The concepts and their implications. *Archivaria, 39*, 5−10. Retrieved from http://archivaria.ca/index.php/archivaria/article/view/12063/13035

Fear, K., & Donaldson, D.R. (2012). Provenance and credibility in scientific data repositories. *Archival Science*, *12*(3), 319-339. doi:10.1007/s10502-012-9172-7

Flanagin, A.J., & Metzger, M. (2008). Digital media and youth: Unparalleled opportunity and unprecedented responsibility. In M.J. Metzger, & A.J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 5−28). Cambridge: MIT Press.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., ... & Treinen, M. (2001). What makes Web sites credible? A report on a large quantitative study. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 61-68). ACM.

Gefen, D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM Sigmis Database*, *33*(3), 38-53. doi:10.1145/569905.569910

Giaretta, D. (2011). *Advanced digital preservation.* Heidelberg: Springer.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis* (7th Edition). Upper Saddle River, NJ: Prentice Hall.

Hilligoss, B., & Rieh, S.Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing and Management*, *44*(4), 1467-1484. doi:10.1016/j.ipm.2007.10.001

International Organization for Standardization. (2012). Space data and information transfer systems: audit and certification of trustworthy digital repositories [Systèmes de transfert des informations et données spatiales: audit et certification des référentiels numériques de confiance] (1st Edition). Geneva, Switzerland: International Organization for Standardization.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141-151. doi:10.1177/001316446002000116

Kaiser, H.F., & Rice, J. (1974). Little jiffy, mark iv. *Educational and Psychological Measurement, 34*(1), 111-117. doi:10.1177/001316447403400115

Kim, Y. (2015). "Designated communities": Through the lens of the Web. *International Journal of Digital Curation, 10*(1), 184-195. doi:10.2218/ijdc.v10i1.360

Kline, R.B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher, C. Schatschneider & D.L. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 171−207). New York: Routledge.

MacNeil, H. (2000). *Trusting records: Legal, historical and diplomatic perspectives*. Dordrecht, London: Kluwer Academic.

Metzger, M.J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078-2091. doi:10.1002/asi.20672

NESTOR Certification Working Group. (2013). *Explanatory notes on the nestor seal for trustworthy digital archives*. Germany: Deutsche Nationalbibliothek.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Rieh, S.Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, *53*(2), 145-161. doi:10.1002/asi.10017

Rieh, S.Y., & Belkin, N.J. (1998). Understanding judgment of information quality and cognitive authority in the WWW. Paper presented at the *ASIS Annual Meeting,* Pittsburgh, PA. Retrieved from http://rieh.people.si.umich.edu/papers/asis98.pdf

Ross, S., & McHugh, A. (2005). Audit and certification of digital repositories: Creating a mandate for the Digital Curation Centre (DCC). *RLG DigiNews*, *9*(5). Retrieved from http://worldcat.org/arcviewer/2/OCC/2009/08/11/H1250012768062/viewer/file2.html

Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651-665.

Rotter, J.B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*(5), 443-453.

Spector, P.E. (1992). Summated rating scale construction: An introduction. Newbury Park, California: SAGE Publications.

St. Jean, B., Rieh, S.Y., Yakel, E., & Markey, K. (2011). Unheard voices: Institutional repository end−users. *College and Research Libraries, 72*(1), 21−42.

Sundar, S.S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M.J. Metzger, & A.J. Flanagin (Eds.), *Digital media, youth, and credibility*. Cambridge: MIT Press.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics.* 4th Edition. Boston: Allyn and Bacon.

Washington State Archives − Digital Archives. (2016). Background and history. Retrieved from http://www.digitalarchives.wa.gov/StaticContent/background

Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority.* Westport, Connecticut: Greenwood Press.

Van House, N.A., Butler, M.H., & Schiff, L.R. (1998). Cooperative knowledge work and practices of trust: Sharing environmental planning data sets. *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*. doi:10.1145/289444.289508

Van House, N.A. (2002). Digital libraries and practices of trust: Networked biodiversity information. *Social Epistemology, 16*(1), 99−114. doi:10.1080/02691720210132833

Van House, N.A. (2003). Digital libraries and collaborative knowledge construction. In A. Peterson-Kemp, N.A. Van House & B.P. Buttenfield (Eds.), *Digital library use: Social practice in design and evaluation*. Cambridge, MA: MIT Press.

Yakel, E. (2004). Seeking information, seeking connections, seeking meaning: Genealogists and family historians. *Information Research, 10*(1) Retrieved from http://InformationR.net/ir/10-1/paper205.html

Yakel, E., Faniel, I., Kriesberg, A., & Yoon, A. (2013). Trust in digital repositories. *International Journal of Digital Curation, 8*(1), 143−156. doi:10.2218/ijdc.v8i1.251